

The Digital Object Identifier initiative: metadata implications

DOI discussion paper Number 2 (version 3, 10 February 1999)

Authors: Norman Paskin (n.paskin@doi.org)
Godfrey Rust (godfreyrust@dds.netkonect.co.uk)

Context/status information

The purpose of this document is to generate a DOI Foundation consensus paper on metadata, as a companion paper to the earlier consensus paper on *DOI Current Position [DOI]*. This paper represents a summary of work in progress and is subject to amendment at any time. *Comments and criticisms* are welcomed. This version has been compiled following comment on earlier versions. Further amended versions may be posted to the DOI web site at <http://www.doi.org> and may include changes of policy or proposed direction in light of comments and criticisms received, which are welcome and should be sent to n.paskin@doi.org or to the DOI discussion mailing list at discuss-doi@doi.org. Nothing in this document should be construed as a commitment to permanent policy of the International DOI Foundation. This paper may be reproduced and distributed for the purposes of comment, providing that it is reproduced in total including this statement

Authors: Norman Paskin, Godfrey Rust*
Date: 10 February 1999
Version: 3
DOI: 10.1000/131

Acknowledgements

*Godfrey Rust (INDECS/Data Definitions) is the main author of section 5 of this paper (detailed DOI metadata elements in relation to INDECS) and the appendix of preliminary examples, and has also made substantial helpful contributions to the concepts throughout this paper.

CONTENTS

1. **Introduction.**
2. **DOI: next steps**
3. **Policy issues**
4. **Underlying approach used to define by DOI metadata elements: well-formed metadata.**
 - 4.1 **The choice of elements which allow interoperability**
 - 4.1.1. **A common key/core/mandatory set of elements guarantees interoperability**
 - 4.1.2 **Well-formed metadata**
 - 4.1.3 **DOI essentials for metadata**
 - 4.1.4 **Metadata need not be centralised**
 - 4.2 **Conformance with existing or developing standards**
 - 4.2.1 **Standards for set of elements: the use of INDECS**
 - 4.2.2 **Standards for means of expressing and interchanging metadata: the use of RDF**
5. **DOI metadata in detail**
 - 5.1 **Outline of the approach used**
 - 5.1.1 **DOI metadata declarations**
 - 5.1.2 **DOI element set**
 - 5.1.3 **Key (mandatory) metadata elements for DOI registration**
 - 5.1.4 **Element qualifiers**
 - 5.1.5 **Key metadata extensions for specific DOI Genres**
 - 5.1.6 **Element values**
 - 5.2 **The DOI metadata elements**
 - 5.2.1 **DOI**
 - 5.2.2 **DOI Genre**
 - 5.2.3 **Identifier**
 - 5.2.4 **Title**
 - 5.2.5 **Descriptor**
 - 5.2.6 **Type**
 - 5.2.7 **Origination**
 - 5.2.8 **Mode**
 - 5.2.9 **Form**
 - 5.2.10 **Extent**
 - 5.2.11 **Context**
 - 5.2.12 **Subject**
 - 5.2.13 **Event**
 - 5.2.14 **Creation Link**

- 5.3 Some issues
 - 5.3.1 Authority issues: Record version number and more
 - 5.3.2 Mapping to other namespaces
 - 5.3.3 Relation to Dublin Core
- 5.4 Summary of benefits
- 6. Building applications of metadata
 - 6.1 IDF role in developing applications
 - 6.2 Relating concepts of applications and functional metadata to the “well-formed” metadata approach
 - 6.3 STM applications: Reference linking
- 7. Possible technical solutions for making metadata declarations and using metadata records
 - 7.1 Multiple resolution
 - 7.2 URN registration schemas.
 - 7.3 RDF schema registration
 - 7.4 DOIs for metadata (and other specified types) – some preliminary comments
 - 7.5 Metadata records and DOIs
- 8. Objections to the suggested approach
 - 8.1 “Isn’t it better to start small?”
 - 8.2 “Why discourage diversity?”
 - 8.3 “It will be too much work for owners”
 - 8.4 “Lets wait and see”.
- 9. Long term implications of DOI metadata: an infrastructure for content in a digital environment
- 10. Implications
- 11. Summary: proposed DOI metadata development strategy

References

APPENDIX: Examples of DOI metadata following the basis set out in the Discussion paper

- Example 1: Book
- Example 2: Music CD
- Example 3: Poem
- Example 4: Journal issue
- Example 5: Journal article

1. Introduction.

The current version (3) of this DOI consensus paper on metadata adds to the previous version much more specific proposals on DOI metadata elements. This forms section 5 of the current paper (which readers of the earlier versions, or those seeking detailed metadata technical discussion, may wish to go to directly), together with the Appendix. It is stressed that Section 5 is proposed, deliberately, as work in progress. It is to be tested in relation to many more practical real-world examples, as part of the DOI-INDECS collaboration to arrive at a definitive stable set and structure which will be formally standardised. However the current proposed set and kernel elements appear to be relatively stable and we think are unlikely to change substantially. Comments and feedback are welcomed. The contents of this paper will be used at several meetings over the next few weeks to elicit responses.

The present paper is necessarily lengthy and detailed. It has to be so in order that the concerns of those experienced in metadata issues, as well as those for whom this is a new topic, can be addressed. When a consensus has been reached and a stable agreed version is available, an abbreviated “executive summary” version will be prepared for use in less detailed contexts. In due course a clear, practical Allocator’s Guide for those issuing metadata related to DOIs will be produced, and it is expected that many of the practical aspects will be managed by available technology.

2. DOI: next steps

The first DOI discussion paper [DOI1] outlined the current position and development path of the DOI. It proposed that limiting the DOI to the initial implementation of one DOI to one URL is too restrictive, and that if such an identifier scheme was to offer sufficient potential to justify its costs it is necessary to:

- Add metadata associated with each DOI entity at the time of its allocation (registration), to enable the construction and implementation of applications of the DOI;
- Add a multiple resolution protocol, which would enable some services to be standardised and fully automated.

This position was endorsed by the Board and is now a commitment of the IDF.

DOI policies today explicitly or implicitly state the following:

- A DOI identifies a Creation, not an address or service;
- DOI may identify any type of Creation (or, as we shall see below, its related metadata document)
- The DOI should be capable of being processed by any software application
- The DOI is a “dumb” identifier

DOI ambitions for the future are:

- DOI will be a kernel component of electronic rights management;
- DOI will be capable of operating for all media in all sectors;

- Most DOI-based applications will be automated
- DOI will fill the role of the unifying managed, quality-assured, digital resource identifier for materials of significant intellectual worth.

3. Policy issues

The role of the IDF in metadata has been of concern to some members of the Foundation. If DOI registration agencies are collecting core (kernel) metadata, will that be used to build applications preventing others from doing so? No: the role of the IDF is “to support the needs of the intellectual property community in the digital environment”, by promoting the DOI as a standardised solution for the intellectual property communities (including text, music, images, and multimedia). The IDF therefore wants to have DOI used in *multiple applications*. It is not the role of the IDF to build *specific* applications of the DOI (although as noted in section 7 below, it will define a set of priority services and genre applications, and may participate in their development even if eventually outsourced to third parties. The DOI kernel registration metadata is intended to be “core”/ “bare bones”/ “minimal”. The Foundation should ensure that DOIs can be used in as wide a range of applications as possible and encourage the development of applications by others. These applications will probably use the core DOI registration metadata but add to it other features (additional metadata, or applications linking to other resources) for specific purposes (including applications in defined Genres, as discussed below).

The DOI can apply to any area of intellectual property. The purposes to which DOIs will be put will be many and various, and unlimited. Therefore the DOI metadata structure must be that which can be used interoperably in any application environment.

IDF must define the business model which should apply to the DOI metadata. The DOI Foundation does not intend to set up a separate business to exploit this data, both because (1) IDF is a not-for-profit organisation intended to set rules and policies, not implementer of DOI-mediated trading; and also because (2) such a business may be in conflict with the business interests of several members. The aim of the Foundation should be to promote the use of DOIs, by ensuring that the rules and policies are clear and equitable. However there are several reasons why a business model is necessary:

- If DOI agencies are collecting metadata, this may be an activity which offers potential economies for others: metadata is collected once (“do it once, do it right”) and made available to those organisations which might want to re-use such data, on some yet-to-be-defined basis. How should the benefits of such economies be realised?
- If one specific agency is tasked, as a registration agency, with collecting and qualifying core metadata, it will incur costs which it may expect to recoup, either directly or indirectly by having some preferential right of usage.
- IDF may ultimately be funded by some mechanism which generates income from DOI applications (unlike now); however these applications will be developed by other organisations. A means of connecting the two functions must be devised.

So questions which arise include:

- what are allowed uses of DOI metadata, by the Foundation, by the Foundation’s agents (e.g.

registration agencies), and by others?

- on what basis or terms should the Foundation make available access to such data?
- what rights should DOI-assigning publishers have to metadata assigned to their entities?
- are there any legal implications of these discussions which we need to consider?
- would there be preferential terms for members?

The policies regarding metadata need to be incorporated into the agreements now being considered for registration agencies. Some members of the IDF have expressed concern that they will become the victims of such a policy; we have therefore encouraged members to formulate the policy themselves. A working group of members has been tasked with drafting an initial policy statement and reviewing areas of concern.

Policy issues are separate from the technical discussion of what the core metadata consists of, but will be influenced by these considerations.

4. Underlying approach used to define by DOI metadata elements: well-formed metadata.

The DOI system must contain some intelligence somewhere, and recent discussions have concluded that the intelligence must be kept out of the identifier itself [NISO-DOI], which will therefore be a “dumb” identifier [Paskin]. Therefore the DOI system must have reliable, predictable metadata. This is not to say that URLs are not useful (dumb pointers which generally contain semantics -which can sometimes be misleading), or that PURLs are not useful (persistent dumb pointers), or that Handles unburdened with DOI policy are not useful (persistent multiple dumb pointers with some typing plus other techniques; but it is to say that the DOI is going to be defined functionally in such a way that there must be associated reliable understandable metadata. That is, the need for metadata should simply and logically fall out of a description of the purpose of the DOI. Given the decision to add metadata, the next decision is, “which pieces of metadata, and within what structure?”

It was noted in IDF Directors report #4 [IDF Dirrep4] that “although it is tempting to define a metadata set for one specific application area, future interoperability issues (e.g. of video clips within scientific articles) must be considered. An incorrect decision at this point could be a significant set back in the future”.

There are two kernel issues which we must bear in mind in this choice:

- The choice of elements which allow *interoperability* between applications;
- Conformance with existing or developing *standards* (for choice of element set, and for means of expressing the set), as has always been the case for our DOI developments.

4.1 The choice of elements which allow interoperability

4.1.1. A common key/core/mandatory set of elements guarantees interoperability

The choice of a structure for elements which allows metadata interoperability is recognised as a key issue in many current discussions such as Dublin Core, INDECS, RDF, AAP's ETC subcommittee on metadata, etc. Other approaches have used standard library cataloguing approaches for one particular, resource discovery, application (e.g. the InterCat experiment: [Jul]).

It is proposed here that the solution lies in which is "well-formed" (application-neutral; structured as logical entity/attributes); and that this well-formed metadata be expressible in RDF.

The intelligence derived from an identifier system must lie with metadata rather than being embedded within intelligent identifiers, if the system is to be extensible and used in many contexts [Metadata].

A given entity to which an identifier is applied may have associated with it, in the identifier system, data which provides additional information, e.g. about its content, rights, etc. This metadata is potentially an infinite set: there is no such thing as "all of the metadata" for an entity as someone may devise a system which uses a piece of associated data not previously considered and recorded in the identifier system. Note also that each of these elements may also be a piece of intellectual content in its own right - the elements of metadata are just as much data as the data is, and may even be part of it

Our goal is therefore twofold: first to find, if possible, an overall framework flexible enough to embrace common forms of descriptive metadata from all genres, and at the same time structured enough to make some sense of it; and secondly to define the minimum "mandatory" elements which are needed to arrive at basic identification and interoperability. In other words, what should be the *form*, and what is the minimum *content*, of DOI metadata?

This minimum set would be available to any application which used metadata. For the purposes of the DOI in this paper we refer to this as a *kernel*, to avoid confusion with Dublin Core etc. It is desirable to define and structure this kernel so as to allow the maximum number of application uses of all or part of it: given five core fields of metadata A,B,C,D and E, someone may wish to use A+B+C for one use (service 1); another A+D+E+C for service 2. Each will want to have the metadata fields be usable without "translation" into the application he is devising. We discuss later that the concept of "Genre" is useful in translating this into extended structured metadata requirements for DOI use.

4.1.2 Well-formed metadata

Each element of metadata is a piece of data, an entity in its own right which may be the central concept in another application: metadata is not a type of data, but a relationship between two pieces of data. For example, a person (entity) may have an element of metadata such as "address"; "address" may in turn be the central field in a geographical information system. Since any data

piece can potentially be used anywhere else, the data must be *well-formed* to allow this, rather than designed to be useful in only one functional application. Well-formed (borrowing from the terminology of XML and similar approaches) implies in this context not only developed on the basis of inherent structure, rather than function, but related in a consistent manner.

An XML file can have 3 types of mark-up [Bryan]:

- 1. An XML processing instruction (declares version of XML being used etc)
- 2. A DTD (document type definition)
- 3. A fully tagged document (“document instance”)

Items 1 and 2 are optional in XML. If all three are present the document is said to be <valid>. If only item 3 is present, the document is said to be well-formed. An XML processor will be able to check that it is following the rules of XML, e.g. each element has a start tag and an end tag. The same approach can be applied to metadata: “well-formed” metadata means that it can be parsed in the sense that the structure of the metadata is readable by any “DOI-compliant” software; “valid” metadata means that the metadata values or tags used are fully meaningful within the DOI namespace (or within a namespace such as Common Information System or Library of Congress recognised as a valid pointer from the DOI namespace). As with XML, this is an important distinction because it allows extensibility; you can make up your own tags, but you can't make up your own syntax.

The term “well-formed” has also been similarly used in the context of standardised data descriptions and storage of these descriptions in data registries [Newton] by NCITS L8 (Data Representation) and ISO/IEC JTC 1/SC14 (now SC32) which resulted in [ISO 11179], part 5 of which covers naming and identification principles for data elements (new work is now underway which includes mapping existing data to data registries and a replacement of part 3 of ISO 11179 with ANS X3.285, a standard which presents the same material and more in terms of a metamodel [NIST]).

Well-formed metadata has the following attributes:

- Values of elements coded according to the rules of a defined *namespace*, or a controlled vocabulary (so a free text description is not normally acceptable, except in the case of necessarily “self-standing” or “autonomous” labels like title), but a standardised code such as Standard Address Number is, since it enables ready re-use in other contexts);
- Structured in relation to other entities by means of a formal data model, with entity/attribute relationships following well-defined rules. This enables any output design to incorporate the entities and attributes into their own model;
- Uses a high-level, generic common data dictionary and data model, from which the set of standardised labels for metadata are chosen, and for which qualifiers may be defined in a systematic way (if the data model and dictionary are defined within the DOI system we might say the metadata becomes fully “valid”, but this is not essential for an appreciation of the more important underlying important concept of “well-formed” data).

The analysis of well-formed or structured metadata and the development of an interoperable set is the task of the INDECS (*Interoperability of Data in E-Commerce Systems*) project [INDECS].

The IDF is an affiliate of the INDECS activity and is using this as a platform to progress its metadata development. INDECS has significant support from a wide range of the “DOI community”; it is working on a tight schedule; it is practically oriented; it is run by experienced and trusted individuals with a proven experience in building practical implementations of metadata. It is important to recognise that INDECS is an example of the well-formed approach used implicitly or explicitly in a number of other metadata activities, notably IFLA’s Functional Requirements for Bibliographic Records [IFLA] and the Common Information System for musical and other works [CIS]. Further, it has reached a consensus with the Dublin Core community on the same approach [Bearman].

The “well-formed” approach is to define entities on the basis of their inherent structure, rather than their application roles (e.g. place, person, creation, right, event, agreement); descriptive elements of entities (e.g. for creations, elements might include the identifier or label itself; controlled vocabulary descriptor labels, event, extent, relation) and types within these elements (e.g. within the element “extent” would be the types “spatial” and “temporal”). One of several examples of existing metadata models following this “well-formed” approach is IFLA’s Functional Requirements for Bibliographic Records [IFLA].

4.1.3 DOI essentials for metadata

From the current policies and future ambitions set out in section 1, and the acceptance of the well-formed metadata principle, we can derive the following:

- Any DOI may turn up in any application;
- Applications must have a predictable answer to the question “what does this DOI identify?”
- The only available point of quality control for this is at the point of allocation;
- DOI key metadata must therefore be:
 - Declared on allocation
 - Built from accessible vocabularies
 - Accessible itself
 - Adequate for identification
 - Predictable to applications
 - Extensible for local needs
 - Persistent
 - Authorised (the question of authority is discussed further below).

4.1.4 Metadata need not be centralised

The Internet creates a situation where metadata may be created once and remain in its original location and yet be accessed by everyone (it is not essential that the metadata be stored in one central location, merely that it be readily accessible and somehow authorised). The Internet Scout Project [Roszkowski] for example is attempting to use distributed metadata for resource

discovery. Equally, we want to minimise the size of the DOI core set and the work needed to compile it (reducing unnecessary duplication). So we want the *input* which creates the kernel to be done once, in a well-defined way (“do it once, do it right”), which does not reflect a specific function but allows multiple functions. This can be done by designing the kernel metadata to reflect the inherent *structure* of entities and metadata, rather than their *function* in any one context or application [Rust]. Since the only available point of quality control for this is at the point of allocation, there is a strong argument that the kernel DOI metadata must be either centrally registered or registered via a pointer to a validated and authorised source of metadata.

4.2 Conformance with existing or developing standards

4.2.1 Standards for set of elements: the use of INDECS

There have been many efforts to establish metadata structures for particular communities; examples include MARC records for libraries [USMARC], Federal Geographic Data Committee for geospatial data [FGDC], Consortium for the Computer Interchange of Museum Information [CIMI] etc. These provide a tool for those communities, but in the absence of a common data model, interoperability between these “silo” applications is very limited and requires the construction of specific bi-lateral “crosswalks” (mapping transformations between the unrelated data models of each silo), with increasing complexity as the number of silos grows [StPierre]. In a digital environment, this approach is no longer satisfactory as an infinite number of permutations can easily be assembled (and need to be): an author of a scientific article, for example, may well wish to insert a video clip, an animation, and a 3-dimensional structure model into his paper and this must be readily accessible without the need to stop and define a specific “crosswalk” to link the video clip metadata to the text metadata; the links must be automated and immediately usable.

One attempt to overcome this problem began by assuming that a common requirement for interoperable metadata would be resource discovery, and has gained significant momentum and recognition: the Dublin Core [DublinCore] initiative has now defined a core set of metadata for the single functional purpose of resource description: a 15-element metadata element set intended to facilitate discovery of electronic resources. There are significant problems with the DC “simple” core set which limit interoperability (e.g. [Baker]) but recent consensus discussions between of the Dublin Core, DOI, INDECS, and bibliographic standards communities (Oct 30/31 1998) have provided a way forward. It now appear that a set of elements capable of interoperable use will be defined by this large community and receive widespread support [Bearman].

This offers for the first time the possibility of an activity which combines the widespread community participation of the Dublin Core with the well-formed structured approach of INDECS and others (e.g. IFLA, CIS). The DOI community is poised to be able to take advantage of this. If this happens, the support of the DOI added to the existing participants would give an overwhelming momentum to such a standard and its likelihood of adoption and success. A stated aim of INDECS is to take its conclusions and proposed structures through the relevant standardisation processes with the aim of these becoming formal recommendations.

Therefore the DOI should use its status as (a) a well-known initiative with strong cross-community support and (b) an affiliate in the INDECS activity, to leverage this to our advantage. Rather than define our own DOI-specific set of standard core metadata we will conform to the emerging INDECS approach and aid in its development. The major problem with this approach is that the INDECS set is still emerging, although an initial data model is now available [INDECS]. A candidate sub-set appropriate to DOI using the same approach, and compatible with its intent, can be proposed (see below, section 4). It should be stressed that this set is intended to be developed in tandem with both the overall INDECS work (INDECS recognise that some of their developments may use DOI as a focal point for the full INDEC activity) and through testing in specific applications. In working with INDECS, DOI is choosing a well-developed and proven model, but one with enough time before standardisation still to change where it is found to be inadequate.

The INDECS model also encompasses the definition of elements for rights metadata (“agreements”) and its integration with descriptive metadata. This work, developing from the successful agreements model of the CIS Plan, is at a less complete stage than that of “creation” metadata. It will take under its umbrella work such as the Stanford Framework for Interoperable Rights Management (FIRM) and proprietary standards from companies working on copyright management software such as DPRL (Digital Property Rights Language). As DOI will in due course require a structure for declaring rights metadata, IDF is also participating in these discussions.

4.2.2 Standards for means of expressing and interchanging metadata: the use of RDF

There is one clear front-runner for expressing metadata, which we should embrace wholeheartedly as it provides an excellent basis for representing well-formed data models, describing structural entity attributes and relationships [Martin]. This is the Resource Description Framework (RDF) which uses XML (Extensible Markup Language, designed primarily as a means of extending generic mark up capability to the Web) as its interchange format [Miller; Bryan]. Both Dublin Core and INDECS have already made a commitment to using RDF as their Web metadata expression framework. The DOI will join them: IDF has recently joined the World Wide Web consortium and will support the development process of RDF and related activities; we will also develop an RDF schema for our candidate metadata sets.

In this way metadata which is well structured may be encoded for use on the web, allowing interoperability of applications. RDF is a specification developed within the W3C Metadata activity [Miller; RDF] and defines a resource as any object that is uniquely identifiable by a URI. The properties associated with resources are identified by property types, and property types have values. Property types express the relationships of values associated with resources; values may be atomic in nature (text string, number) or other resources (which then in turn may have their own properties). This triadic model of resources, property types, and values is an abstract *data model* which enables an unambiguous method of expressing semantics (thus intelligence); it simply renames the standard data model triad of entities, attributes, and attribute values.

The Resource Description Framework (RDF) Model and Syntax Specification [RDFmod] has been released Jan 7 1999 as a Proposed Recommendation and the W3C Advisory Committee is currently reviewing it. The introduction to this document provides a goof overview of RDF aims: (From the introduction): “The World Wide Web was originally built for human consumption, and although everything on it is machine-readable, this data is not machine-understandable. It is very hard to automate anything on the Web, and because of the volume of information the Web contains, it is not possible to manage it manually. The solution proposed here is to use metadata to describe the data contained on the Web. Metadata is "data about data" (for example, a library catalog is metadata, since it describes publications) or specifically in the context of this specification "data describing Web resources". The distinction between "data" and "metadata" is not an absolute one; it is a distinction created primarily by a particular application, and many times the same resource will be interpreted in both ways simultaneously. Resource Description Framework (RDF) is a foundation for processing metadata; it provides interoperability between applications that exchange machine-understandable information on the Web. RDF emphasizes facilities to enable automated processing of Web resources. RDF can be used in a variety of application areas; for example: in resource discovery to provide better search engine capabilities, in cataloging for describing the content and content relationships available at a particular Web site, page, or digital library, by intelligent software agents to facilitate knowledge sharing and exchange, in content rating, in describing collections of pages that represent a single logical "document", for describing intellectual property rights of Web pages, and for expressing the privacy preferences of a user as well as the privacy policies of a Web site. RDF with digital signatures will be key to building the "Web of Trust" for electronic commerce, collaboration, and other applications”.

5. DOI metadata in detail

This section uses terminology from the INDECS (Interoperability of Data in E-commerce Systems) schema and has been written following extensive discussion between the DOI and INDECS initiatives.

5.1 Outline of the approach used

5.1.1 DOI metadata declarations

Each *DOI* allocated is accompanied by a **declaration** of descriptive metadata. This declaration is made within the overall framework of the DOI element set. The means and location of declarations will be dealt with elsewhere but will include XML, RDF, and registries. DOIs will also need to be accompanied by a declaration of rights metadata: also to be dealt with elsewhere.

5.1.2 DOI element set

The proposed full DOI element set for metadata for any creation contains fourteen high level **elements**. These elements are a specific application of the INDECS generic data model. Twelve

elements are direct attributes of the creation, one is, a related entity, an *Event* in the making or use of the creation, and the last is a link to another creation

5.1.3 Key (mandatory) metadata elements for DOI registration

Of these fourteen, seven form the **kernel** (mandatory) **metadata set**, for each of which at least one value must be declared when a DOI is allocated. Other element values, or additional values of the key elements, may be included on the original declaration, or may be added at any time subsequently.

5.1.4 Element qualifiers

DOI has a set of **element qualifiers**, also derived from the INDECS model, which may be substituted for the highest level elements. These may be used in their own right in declarations, but principally they serve as a means of mapping more closely the elements and qualifiers used in other **namespaces** which may be used freely in DOI metadata declarations. The ability to qualify elements at any level is a major feature of the INDECS schema, and it allows two qualifiers in the *DOI* namespace (that is, *DOI* and *DOI Genre*) to be treated as separate high level elements for *DOI* purposes while remaining entirely compliant with the INDECS schema.

5.1.5 Key metadata extensions for specific DOIGenres

The *DOIGenre* element may determine that additional metadata elements are required for a specific type of creation, as determined by an appropriate interest group and agreed within the IDF. [see *DOIGenre* below for further explanation of this key element]. Because this is linked directly to a controlled value of the *DOIGenre*, extended kernel sets can be specified and validated automatically. *DOIGenres* themselves may be overlapping or hierarchical. A creation falling into more than one *DOIGenre* will be required to meet the extension requirements of each genre. The IDF intends to designate a small number of specific *DOIGenres* and to mandate key extensions for these.

5.1.6 Element values

Values can be taken from any other namespace. For certain elements the DOI has its own namespace values, and in certain specified cases (for example, *Primary Type* and *DOIGenre*) one of these DOI namespace values must be used.

Table 1: Summary of Declared DOI Elements

Key = One value for each *key element* (marked “yes” in the table below) is mandatory.

Number = number of allowed values in total.

Details of DOI/INDECS qualifiers and values, where appropriate, are given in the notes that follow.

| | DOI Elements | Definition | Key | Number |
|----|---|---|-------------------|--------|
| 1 | <i>DOI</i> | A DOI unique to the creation. | yes | 1 |
| 2 | <i>DOIGenre</i> | A homogenous category of creations drawing on common namespaces and/or key metadata extensions. | yes | Many |
| 3 | Identifier | A unique label allocated to a creation by a designated authority. | yes | Many |
| 4 | Title | A name by which the creation is known. | yes | Many |
| 5 | Descriptor | A description applied to a creation. | | Many |
| 6 | Type | The basic structural type of the creation. | yes | 1 |
| 7 | Origination | A process by which the creation was made. | yes | Many |
| 8 | Mode | A sensory mode through which a creation is perceived. | | Many |
| 9 | Form | A format or genre of the creation. | | Many |
| 10 | Extent | A measurement of the form of the creation. | | Many |
| 11 | Context | A purpose, audience or environment for which or within which a creation was made. | | Many |
| 12 | Subject | An external element which is a subject of the content of a creation. | | Many |
| 13 | Event Type Agent Role Tool Time Place | Something that happens to a creation | yes yes yes | Many |
| 14 | Creation Link | A link to another creation. | | Many |

5.2 The DOI metadata elements

5.2.1 DOI

A DOI unique to the creation.

Key element: one *DOI* must be declared.

5.2.2 DOI Genre

A homogenous category of creations drawing on common namespaces and/or key metadata extensions.

Key element: at least one *DOIGenre* must be declared. The *DOIGenre* is the “Rosetta Stone” element which enables specific interest groups to develop their own applications and metadata schemas, while remaining wholly consistent and interoperable with the universe of DOIs.

DOIGenres will coincide with recognised product groups and/or market which commonly have their own identifier systems, metadata namespaces and interest groups and trade associations, but which cannot, in the multimedia digital environment, be strictly partitioned. There might be *DIOGenres* such as *book*, *journal*, *audio* or *audiovisual*. They look simple at first glance, but in reality are complex and overlapping categories defined by combinations of two or more INDECS attributes, typically *PhysicalForm*, *Genre*, *Mode* and *Audience*.

The *DOIGenre* element allows the needs of these sectors to be met without disrupting the integrity of the DOI system as a whole. For example, a scientific journal with a DOI must carry the necessary metadata and values to enable the DOI to support academic citations (whatever that community may decide them to be) while remaining able to function consistently with applications handling DOIs of any kind for discovery, licensing, retailing and royalty distribution in the wider digital environment (see further discussion under section 6.4).

Each scientific journal article must therefore declare the *DOI* kernel metadata elements, plus whatever additional elements from within the full element set are determined by the *DOIGenre* (say, *journal*) by which it is categorised. These **kernel metadata extensions** will be determined by appropriate interest groups within IDF, with reference to sector and identifier schemas.

The *DOIGenre* may also determine specific namespace values required for describing that particular creation’s metadata. For example, the *DOIGenre* of *journal* may require that at least one *Identifier* from the types ISBN, ISSN or CODEN be used; or the *DOIGenre* for *book* may require that values from the BIC/BASIC/EDItEUR Data Dictionary set be used. Any of these will be compatible with *DOIs* as they are being mapped back to the DOI element set using the INDECS schema (see *Mapping to other namespaces* below).

DOIGenres may be overlapping. For example, a spoken-word audio product might come in *DOIGenres* for both *Audio* and *Books*, and would therefore require the extended key metadata elements for both. *DOIGenres* may also be hierarchical. For example, *Childrens Books* might be a *DOIGenre* which is a sub-type of *Books*, for which the *Audience* class becomes a mandatory element.

The identity of *DOIGenres* requires definition (a work item for the DOI Foundation, to be dealt with in a separate paper).

5.2.3 Identifier

A unique label allocated to a creation by a designated authority.

Key element: at least one *Identifier* other than the *DOI* must be declared. This may be another public or standard identifier (ISBN, UPC, ISRC, SICI etc), or it may be the declarer's own internal reference by which the item is identified within his own system. The namespace (designated authority) of the identifier must be declared. This ensures interoperability with existing, "legacy" identifiers (e.g. ISBN, ISSN, ISRC, etc.).

5.2.4 Title

A name by which the creation is known.

Key element: at least one *Title* must be declared.

5.2.5 Descriptor

A description applied to a creation.

Optional. *Descriptors* contain free text of any length. There are three INDECS/*DOI* qualifiers to which namespace qualifiers may be mapped:

INDECS qualifiers for **Descriptor**

| | |
|-------------|--|
| Keyword | A word or phrase associated with a creation, taken from a controlled or uncontrolled vocabulary, |
| Description | A textual description of a creation |
| Annotation | An explanatory note about a creation |

5.2.6 Type

The basic structural type of the creation.

Key element: one value only must be declared from the *DOI* namespace set below. This is the 'Primary Type' which distinguishes between the fundamental types of creation. Within the INDECS scheme there are three values (Manifestation, Expression and Work). However, in the *DOI* element set these are qualified somewhat at the highest level: Manifestations are subdivided into *Physical* and *Digital*; and Expressions are recognised only in the form of *Recorded Performances*.

DOI namespace values for **Type**

| | |
|------------------------|--|
| Physical Manifestation | A creation realised in physical form (“atoms”) |
| Digital Manifestation | A creation realised in digital form (“bits”) |
| Recorded Performance | Also known as Recording. A creation realised in space and time (“actions”) and encoded within a physical or digital manifestation in a manner that allows for its recreation |
| Work | A creation of the mind whose existence is revealed through an expression or manifestation |

5.2.7 Origination

The process by which the creation was made.

Key element: all valid values be declared from the DOI namespace set below. This element determines whether the creation is original or is derived in some way from other creations of the same type. This is essential for determining whether source or component creations need to be identified for rights management purposes. In some *DOIGenres* (for example, Musical Works) this value will determine whether information on *Creation Links* (element 14) become key metadata. Creations can have multiple values of Origination: in extreme cases a creation may simultaneously be a version, a compilation and an excerpt.

DOI namespace values for **Origination**

| | |
|-------------|--|
| Original | A creation whose existence is not recognisably dependent on a pre-existing creation of its own type |
| Version | A creation made through the transformation of a pre-existing creation of its own type (= Modification, Transformation) |
| Compilation | A creation made through the compilation of two or more pre-existing creations of its own type |
| Extract | A creation made through extracting a part of a pre-existing creation of its own type (= Excerpt) |
| Replica | A creation made by copying a pre-existing creation (= Copy) |

5.2.8 Mode

A sensory mode through which a creation is perceived.

Optional. This element helps to determine which classes of actions and rights are appropriate to a given creation. Three of these (relating to touch, taste and smell) will have little relevance to most entertainment, reference or educational content providers, but have potentially useful

applications in identifying creations subject to rights and patents in industries such as fashion, perfume and food. Modes can be used in any combination, so the rights-based characteristics, for example, of a fashion item would be both *Visual* and *Tactile*, and of an audio both *Audio* and *Visual* (the latter relating to the cover design, sleeve notes etc.). There is an intentional anomaly in this list with the inclusion of *Audiovisual* as a single *Mode*, to distinguish moving from static visual images.

*DOI namespace values for **Mode***

| | |
|-------------|--|
| Audio | A creation perceived through the sense of hearing |
| Visual | A creation perceived through the sense of sight (static images) |
| Audiovisual | A creation perceived simultaneously through the sense of hearing and sight (moving images) |
| Tactile | A creation perceived through the sense of touch |
| Gustatory | A creation perceived through the sense of taste |
| Olfactory | A creation perceived through the sense of smell |
| Abstract | A creation perceived only by the mind, by abstraction from a related expression or manifestation |

5.2.9 Form

A format or genre of the creation.

Optional. The *DOIGenre* and *Mode* already provide basic formal information. However, this element embraces a myriad of metadata of importance for many purposes, so it is likely that every extended key metadata set will include some mandatory requirements for declarations of form which will vary from *DOIGenre* to *DOIGenre*. For example, a digital document may require a *MIME* type, a sculpture may require a declaration of *Material*, or a production may require an *Audiovisual Genre*.

Form has an almost limitless number of possible namespace qualifiers and values. The seven INDECS generic qualifiers below are designed to provide comprehensive mapping for the distinctive types of form metadata. Analysis of the list shows that the first four qualifiers relate to manifestations, and the latter three to expressions. As manifestations contain expressions, the latter elements can be applied to manifestations “by proxy”: so, for example, a *Language* value attached to a book is interpreted as meaning “the Language in which the content of this manifestation is expressed”.

None directly relate to abstract works, because abstract works have no definitive form. However, any of these values can be applied to abstract works in the sense of being forms that they are “normally expressed/manifested in”.

INDECS qualifiers for **Form**

| | |
|-----------------|---|
| Physical Format | The format of a physical manifestation |
| Digital Format | The format of a digital manifestation |
| Material | The material of a manifestation in or on which an expression is recorded |
| Encoding | The technique of representation or encoding that fixes an expression in or on a manifestation |
| Genre | A class in which the form of expression is recognised |
| Language | The language of expression |
| Completion | The status of completeness of expression |

5.2.10 Extent

A measurement of the form of the creation.

Optional. Extents are complex elements in that they have three parts: a qualifier or *Extent Type* (for example, *Height*, *Capacity*, *Duration* or *Number of Pages*), a quantity (the number itself), and an *Extent Measure* (for example, *mm*, *MegaBytes*, *HHMMSS* or *units*). Every *Extent* must have all three. INDECS identifies three high level qualifiers:

INDECS qualifiers for **Extent**

| | |
|-----------|---|
| Dimension | A measurement of spatial characteristics of a creation |
| Duration | A measurement of temporal characteristics of a creation |
| Number | An enumeration of a component of a creation |

5.2.11 Context

A purpose, audience or environment for which or within which a creation was made.

Optional. Context provides for characterising external relationships of function and purpose, especially relevant for identifying target or appropriate audiences for a creation.

INDECS qualifiers for **Context**

| | |
|-------------|---|
| Audience | An intended audience for a creation |
| Purpose | A purpose for which a creation was made |
| Environment | An environment within which a creation was made |

5.2.12 Subject

An external element which is a subject of the content of a creation.

Optional. There are more extensive namespace value lists for subject than for anything else (for example, Dewey and Library of Congress). INDECS, in harmony with the IFLA *Functional Requirements for Bibliographic Records*, identifies six high-level qualifiers. Like *Genre*, *Language* and *Completion* already mentioned, *Subject* is another “proxy” element which applies to the content (the expression or abstract work) but is attached to the manifestation.

INDECS qualifiers for **Subject**

| | |
|---------------|--|
| Topic/Concept | A topic which is a subject of the content of a creation |
| Person | A person or class of persons which is a subject of the content of a creation |
| Creation | A creation or class of creations which is a subject of the content of a creation |
| Thing/Object | A thing or class of things which is a subject of the content of a creation |
| Event | An event or class of events which is a subject of the content of a creation |
| Place | A place which is a subject of the content of a creation |

5.2.13 Event

Something that happens to a creation.

Key metadata: every *Event* involving a primary creator must be declared. *Events* cover the making and use of creations, and this element requires some special consideration. It is an efficient, flexible way of recording all acts of “making” or “using” creations. The conventional way of recording elements such as the identity and role of the author, date of creation, publisher, territory of publication and other related elements are as simple attributes of the creation itself: “Author=John Brown”, “Date of Composition=1965”, Place of Composition=London” etc. This leads to complications when creations are distributed in more than one territory, or by two publishers, or when several people (known here as *agents*) contribute at different times and places and using different tools to the creative act. It is impossible in such a simplistic model to accurately describe each individual contribution, and if those specific contributions are subject to rights, this is likely to become a problem for any metadata supporting rights management.

For a simple example, when Jill Smith sets the 1965 poem of John Brown’s to music in Philadelphia in 1985, the traditional approach has a problem. The new work will have two creators (Author=John Brown, Composer=Jill Smith), and at best two dates of composition

(=1965, 1985) and two places of composition (London, Philadelphia), but no structured way of linking them.

The DOI/INDECS *Event* groups together the agent, role, tool (if any) and the time and place at which the creative or distributive act took place. In the example quoted, there are two creative events: *John Brown, Author, 1965, London* and *Jill Smith, Composer, 1985, Philadelphia*. This approach allows complete flexibility to be as broad or narrow as necessary, and XML-based applications will allow the “nesting” of events to give multiple roles to agents, or group agents together in time and/or place.

In the key set, only the *Agent* and *Role* elements are mandatory, so at minimum the declaration is as simple as under a traditional approach (Author=John Brown), except that when required any combination of elements can be declared accurately. All primary creators must be declared (according to “primary” categories of creator role determined within the *DOI*Genre).

Many Creations have multiple authors: the average number of authors per paper in the 1997 ISI database is 3.84 (and the trend is rising (source: ISI, pers. comm.); in music, the average number of composers is 1.9 and also rising (Rust, pers.comm.). Extremes such as 40-author physics papers are not unknown. All primary creators must be declared, even if there are 40 of them. For discovery or rights it is essential. It has to be all or nothing: declaring only some of a set is critically dangerous for E-commerce (imagine the consequences of declaring only Lennon as the writer of the Beatles' songs).

“An event is, of course, a distinct entity from the creation itself, with its own attribute set. Events may be set at any level of granularity. For example, the making of an audiovisual production will involve many people in many roles employing many tools at different times and places to produce a number of different creations. This may be described as a single event in a broad time period (say January-March 1998) and place (England) in which a large number of people did different things. Or it may be broken down to specific contributions, so a particular actor may have acted in a particular scene on a given day in Winchester. How granular events need to be will vary from *DOI*Genre to *DOI*Genre and from creation to creation, following the INDECS principle of *functional granularity*. If it is recognised to be of commercial or cultural value, events may be separated out. An article with twenty authors may simply need one event for all, for only their identity and not their precise contribution or its location may have any significance; while a sound recording in Europe and the USA will need to identify the location of each separate recorded contribution in order to determine eligibility for royalties under European laws.

In the INDECS schema *Events* are of three types: *Making*, *Dissemination* and *Use*, and this enables all events in the life of a creation to be documented.

DOI/INDECS attributes for **Event**

| | |
|------------|---|
| Event Type | The type of event (Creation, Dissemination, Use) |
| Agent ID | The identity of a person playing a role in the event |
| Agent Role | A part played by a person in an event |
| Agent Tool | A thing used by a person while playing a role in an event |

| | |
|-------------|--|
| Event Time | The time during which the event occurred |
| Event Place | The place in which the event occurred |

The term “Publisher” is a common source of confusion in metadata. The INDECS analysis enables us to separate three distinct ways in which the term *Publisher* is commonly used. Where a publisher is responsible for commissioning and/or producing a creation, he plays an agent role as Publisher in a *Creation Event*. Where a publisher is responsible for making something available to the public, he plays an agent role as Publisher in a *Dissemination Event*. The third meaning (common among music publishers) is that of Rights Owner, which does not belong here but would be included in a *Rights Metadata* declaration. The DOI element set therefore allows Publishers to simply and accurately describe their roles in relation to each creation. For many DOIGenres we expect that the role of Publisher would be a key element.

5.2.14 Creation Link

A link to another creation.

Optional. Creations may be linked to one another in three ways: as a *Component* (for example, a track on a CD, a poem in an anthology or a film clip), a *Version* (for example an edition of a book, remix of a recording, translation of a text) or a *Reference* (for example, a journal citation). These links may be declared in a standard form independent of the original metadata declaration.

Creation links, particularly component creation links, are likely to be among the most common connectors in the digital metadata network, joining together declarations produced by different parties for different DOIs to represent complex creations. For example, an edition of the novel *1984* would, when its declaration is complete, comprise one declaration for the physical manifestation, one for George Orwell’s abstract work, and one for the component creation link between them.

For a more complex example, a journal containing fifteen articles, three of which contained respectively two, five and three diagrams, would when complete comprise 1 declaration (manifestation) with fifteen links to fifteen article declarations (works) which in turn would have links where appropriate to the ten diagrams (manifestations).

Of course, the articles and diagrams need not be declared as part of the key metadata, and may in practice never need to be declared at all if the authors have assigned all rights to the publisher or have no interest in or means of benefiting from further exploitation. Equally it may only be a few of the articles which require declaration. But if any or all of them require identification, the DOI/INDECS schema enables it at whatever level of detail or granularity is required, no more and no less.

DOI/INDECS attributes for **Creation Link**

| | |
|-------------|---|
| Link Type | The type of link (Component, Version, Reference) |
| Sequence | A number determining the order in sequence, if any, in which this link appears among other links of the same type |
| Time | The time at which the creation link was established |
| Place | The place in which the creation link was established |
| Realisation | The technical means used to establish the link |

DOI/INDECS attributes for each creation in **Creation Link**

| | |
|----------------|---|
| Creation 1 DOI | The DOI of one of the two linked creations |
| Extent | The quantity of the creation involved in the link |
| Start Location | The start location of the part of the creation involved in the link |
| End Location | The end location of the part of the creation involved in the link |

Creation Link, describing a Relation between Creations, is not specified as a key (mandatory) item as it will not always (a) exist or (b) be available. However it will be a very important element in certain applications – and therefore DOI Genres (e.g. in specifying, for an electronic journal article, what is the printed equivalent version and the underlying parent work - PII or future ISWC). Like all the other non-key elements it can be included as an option. It is important to note that a wider range of elements can go into a metadata declaration for specific applications, as discussed below - e.g. the full INDECS element set]. If STM publishers want to make it mandatory (for example) in cases where a journal article is previously or simultaneously published, then that would be a local DOI Genre extension, the specific conditions for which we will elaborate as part of such a development.

5.3 Some issues

5.3.1 Authority issues: Record version number and more

Record version number is not in itself a piece of primary metadata but a control in case of additions to or changes to the metadata record: metadata about metadata. Later versions overwrite earlier versions. This is an area for further investigation as it raises the very important issues of authority and security. If automated commercial applications are to be build on DOIs and their associated metadata, these applications will need to be able to trust the metadata that is declared; and for discovery purposes we need to be sure which version, and whose version, we are consulting. What may be required is an "authority" record for each metadata declaration which would include:

- (1) Identity(s) of metadata author
- (2) Version number
- (3) Date

Even this is too simplistic and assumes that the declaration is viewed as a single document. Because people can add their own individual elements to the metadata (for example, I wish to add my own local subject or genre classifications to the "master" record) then each element requires a namespace author and date.

RDF can accommodate this, and it is not a great problem in a structured environment, but it begs the question about how DOI metadata (indeed, any metadata) is going to be maintained: "ownership" means not just authorship, but ongoing control. A technical solution for declarations (based probably on RDF and Handle) needs to be a basic part of the DOI system. If metadata is to be distributed and relied upon, an audit trail is needed to enable updates on an ongoing indefinite basis (of course, rights management requires this to a much greater extent). Metadata declared today may still be operational in twenty years time, linked to hundreds of other metadata declarations and referenced in thousands, even millions, of transactions and licences.

It is recognised that INDECS needs to deal with the issue of authority of metadata ("who says" that this is the correct piece of metadata?). We may momentarily enter an Alice in Wonderland world of absurdity where every metadata record has its own authorising metadata record, which in turn has its own, and so on in infinitely dependent chains. Of course this is ludicrous, and compact technical solutions should be found, but we should note that however extraordinary, these are practical problems that have to be resolved in a global, distributed digital environment. IDF will participate in INDECS' efforts to elaborate this issue.

5.3.2 Mapping to other namespaces

Through the mapping of namespaces being carried out within the INDECS project, and ultimately through a DOI namespace registry (to be developed, it is hoped, in conjunction with the Dublin Core initiative), DOI allocators are able to use metadata created under recognised schemes. This is important in the use of "legacy" systems, in accordance with one of the main tenets of URN identifiers (the ability to interoperate with other conforming identifier schemes).

Through mapping it should prove practical to generate DOI key metadata declarations from any recognised namespace (and vice versa) with a high degree of automation. However, where there are structural weaknesses in the existing namespace some degree of intervention may be required.

5.3.3 Relation to Dublin Core

In view of the importance attached to the Dublin Core effort by many DOI participants, it is useful to consider the relationship of the proposed DOI scheme to this. The fifteen Dublin Core simple elements can be mapped approximately to DOI declarations, but some of Dublin Core's current structural weaknesses cause serious difficulties and gaps, as can be seen in the mapping below. The mapping does not, of course, take account of any standard qualifiers which may be introduced in DC Qualified, and which may bring about some closer correspondence at that level.

However, it is hoped that collaborative work between Dublin Core, IDF and INDECS will ensure that a new expression of Dublin Core rooted in an entity model common to INDECS will bring about much closer, if not complete, integrity between the systems [Bearman].

The DC Element 15 (Rights) is not shown on this map as it corresponds to the as yet undeveloped DOI Rights Metadata Declaration.

| | <i>DOI Elements</i> | DOI Qualifiers | DC Elements |
|----|---------------------|--|--|
| 1 | <i>DOI</i> | | 10 Identifier |
| 2 | <i>DOI</i> Genre | | 8 Type |
| 3 | Identifier | | 10 Identifier |
| 4 | Title | | 1 Title |
| 5 | Descriptor | Keywords Description Annotation | 3 Subject/Keywords 4 Description |
| 6 | Primary Type | | 8 Type |
| 7 | Origination | | |
| 8 | Mode | | |
| 9 | Form | Physical Format Digital Format Material Encoding Genre Language Completion | 9 Format 9 Format 9 Format 8 Type 12 Language |
| 10 | Extent | Dimension Duration Number | |
| 11 | Context | Audience Purpose Environment | |
| 12 | Subject | Topic Person Creation Thing Event Place | 3 Subject/Keywords 14 Coverage? 14 Coverage? |

| | | | |
|----|---|-----------------------------------|---|
| 13 | Event Type Agent Role Tool Time Place | | 2,5,6/Creator/Contributor/ Publisher 7 Date |
| 14 | Creation Link | Component Version Reference | 11,13/Source/Relation |

5.4 Summary of benefits

The proposed DOI metadata approach has the following benefits:

- “Well-formed” key metadata set: entirely predictable for all applications
- Covers all creation types
- Local extensions possible: no constraints on individual sector requirements
- Requires discipline on DOI allocation
- DOI benefits from INDECS/IFLA/CIS/DC (etc) infrastructure work and vice versa
- Allows entry to formal standards process; compatible with current activities such as DC, RDF and URN registration

6. Building applications of metadata

6.1 IDF role in developing applications

To ensure a speedy development of services and applications responding to DOI users’ needs and to promote the DOI system, it would be very useful to define a set of priority services and “killer” applications, even if it will eventually be our policy to outsource their development to third parties. This should in no way hamper or delay consensus on metadata, but would add credibility and understanding of the DOI system and could facilitate outreach to organisations with competence in developing these applications. IDF will also work on defining some key *DOIGenres* and work with appropriate organisations (existing namespace systems, and/or potential DOI registration agencies) to produce approved key metadata extensions for each *DOIGenre*.

6.2 Relating concepts of applications and functional metadata to the “well-formed” metadata approach

It has been suggested (e.g. at the Dublin Core Warwick Framework workshop, by Bill Arms) that metadata can be thought of in functional packages, such as:

- Citation package: author, title, journal, date, issue
- Descriptive package
- Business package (ordering information etc)
- Structural metadata package (formats, plug-ins, protocols etc.)
- Music package
- etc.

These are uses to which selected sets of well-formed (or valid) metadata elements can be put. Not everybody will want to create all packages; it would help enormously if one could ask the DOI Handle “what packages are available for the object with a specified DOI for a user with my permissions?” The IDF guidelines for a registration agency in one particular area (*DOIGenre*) could specify that information objects of a given type should have certain packages. Other organisations could also have guidelines; (e.g. a secondary publisher could state that their systems will preferentially reference materials with a specific package). This would give publishers the choice of providing their own metadata or having it provided by third parties (e.g. Library of Congress and OCLC provide MARC metadata for monographs and other materials; this is equivalent to printing a CIP record in a book.)

By having a small core of well-formed mandatory DOI metadata, it would be possible for different communities (such as e.g. the STM community) to work with their appropriate DOI registration agency or DOI users (like, potentially, ISI) to build additional well-formed or valid metadata elements for a community application (a DOI Genre) yet retain the key interoperability so that when a “non-standard” item like a video-clip is encountered it can be pointed to interoperably within the DOI framework; and that other applications built elsewhere can interoperate with the STM community application

It is likely that this topic will also be dealt with in future discussion papers on services [DOI3] and local resolution [DOI4].

6.3 STM applications: Reference linking

A recent survey has pointed out that the links made by scholars using Internet tools are much more than just citations, including for example other types of “invocations” such as references to abstracts, home pages, list servers, etc. [Cronin]. The same paper quotes proposals that the inherent scalability of the Web could result in the emergence of a universal bibliographic and citation database linking every work to every other work; without structured validation and control of the metadata used to do this, this scenario is not a utopia but a dystopia, where

resource discovery tools process vast quantities of data and are unable to make matches due to different vocabularies, descriptions etc. Reference linking (or the wider concept of matching “invoking” citations) has been discussed within the DOI community because it would be a clear application of the DOI which would offer some aspects of a “killer application” for DOIs and stimulate growth in registration of DOIs. There is a demonstrated need for such linking [Hitchcock; Hunter]; there is a need for “actionable identifiers” [Lynch] which can be used as active links, not just printed citations.

It is clear that this is the main application that many members of the Foundation would put into place immediately. Most are technologically ready to do so. There are existing arrangements such as Web of Science and APS Link Manager which process metadata (citations) to create links; even for small publishers who don’t have (or want) direct agreements there is likely to be a facility such as NLM’s proposed PubRef service which offers the same capability, and which can probably be done by providing some minimal set of metadata for this specific service.

But there are several problems with existing systems which a DOI-based approach solves, which are apparent from the approach proposed here and especially the discussion of Creation links (5.2.14):

1. Existing link creators use URLs, and that these are not sufficiently reliable since surveys have shown typically 50% of e-journal URLs are inoperative [Ford/Harter]. However if the existing or proposed reference linkage mechanisms could be encouraged to adopt DOIs as the identifier for location, this would be beneficial to both.
2. Many STM publishers are already doing linking with other major players, by bilateral agreements. Bilateral agreements grow dramatically more complex to administer as the required degree of coverage increases. DOI offers a much more efficient way of doing such linking.
3. Difficulty of knowing what the citation “really is”. DOI offers the potential ability to solve the issues which all of these services are currently having great difficulty with: that of the “multi-dimensionality” of a scientific article. No longer is it sufficient to assume that a citation to a single printed paper equates to a citation to the work: am I citing the pre-publication released article, the final printed article, the HTML equivalent, or the abstract work, etc.; to do this, it is essential to have appropriate metadata.
4. Difficulty of capturing appropriate data and making efficient use of this by ensuring its interoperability. DOI/INDECS approach is based on the fundamental principle that each element of metadata is well-structured. This provides an easy route to capturing the required metadata in automated processes, since for each element of metadata there is a controlled vocabulary or list of defined options to choose from, such as file format. The exceptions here will be (i) “autonomous” elements which do not refer to a namespace, such as Title; (ii) elements where at present there is no controlled vocabulary (e.g. person identifiers), the potential for upgrading to this once such systems are in place (the INDECS participants are working towards standard declarations for People, Creations, and Agreements in compatible form, and are willing to develop key sets for DOI as a focal point for the full INDECS set.)

Reference linking is a specific application but one which is key to a particular genre, the scientific journal article (or more generally, Cronin et al’s “invocation”). It is therefore appropriate to

design the extended core set for DOI metadata of scientific journal materials so as to offer this application in an easily automated way. (At the time of writing a NISO-sponsored meeting on the topic of reference linking is about to take place, in which IDF is participating, and this may lead to an expansion and revision of this section in future versions of this draft paper.)

The STM application – or specific scientific journal DOIGenre – will be one priority focus of the IDF's efforts to develop applications as soon as possible, based on agreement of a common metadata approach. Other applications in other genres are under consideration.

7. Possible technical solutions for making metadata declarations and using metadata records

Note that:

1. This section outlines some issues which will be dealt with much more fully in the forthcoming DOI discussion paper number 3 [DOI3] on “The Digital Object Identifier: creating services using DOIs” which will take as its starting point the use which is made of the DOI metadata by means of services and multiple resolution schemes; and forthcoming DOI discussion paper number 4 on “local resolution” i.e. how intermediaries (in the widest sense, including libraries and local collections) can make use of DOIs within their services.
2. The possible technical solutions using XML, URN, and RDF mechanisms are the subject of a detailed meeting in mid-February 1999 which will result in expansion and a further iteration of this section.

Although the ideas are still being worked on it is useful to set out some preliminary thoughts for comment. One or many of these might be appropriate and may converge.

7.1 Multiple resolution

Using the DOI's underlying Handle technology. We have defined a mechanism and syntax [NISO-DOI] for “level 1” DOIs (which have a single point of resolution, to a URL) and have outlined [DOI1] the potential for “level 2” DOIs which would have multiple resolution points (both multiple instances of data types such as URLs, and multiple data types). [DOI3] will give examples of use of the existing Handle technology for these purposes. It is easy to imagine the use of a level 2 DOI with a defined data type (either a new type, or possibly an existing type, perhaps even itself a DOI) as a pointer to the key metadata record of the DOI. This also allows one to envisage how metadata records might be distributed: the pointer may be to a registration service, or may itself split the metadata in a defined way among other pointers to other authorised services.

7.2 URN registration schemas.

DOIs can be considered to be implementations of the URN syntax; mechanisms for URN registration are currently under discussion (and IDF is participating in these). These include

possible use of DOIs in collaboration with National Libraries, National Bibliographic Numbers, etc., recently discussed at a meeting in Helsinki between representatives of IRTF/URN, DOI, ISBN, ISSN, and National Libraries.

7.3 RDF schema registration

RDF itself is expressed as an XML schema (DTD) and there will need to be some sort of RDF schema or type registry; that is, we will know how to read and parse RDF but will then have to make sense somehow of the resulting intelligence that this piece of RDF is laying out specifics in terms of the INDECS or other schema. IDF wishes to use RDF as its metadata expression tool. Further details of RDF registration issues will be added to this paper as they become known.

7.4 DOIs for metadata (and other specified types) – some preliminary comments

- Level 2 DOIs will need to be clearly differentiated from level 1 DOIs.
- A metadata record will often be a tradeable item which is itself located by a DOI
- It might be useful to distinguish DOIs allocated to the different primary types: manifestations (digital objects or physical objects, retrievable or accessible via a DOI) as opposed to those allocated to abstract works or performances (not accessible by a DOI, only as meta-objects) or allocated to a metadata record

It would be possible to achieve all these ends by defining the level 2 syntax as incorporating some minimal intelligence relating to “type”: a format such as

A level 1 DOI: 10.1000/123456

A level 2 DOI: 10.1000/W/456789 where W indicates a work (etc)

The inclusion of the additional “type” code delimited by // would indicate a DOI of level 2.

These are preliminary thoughts as yet. A work item for the DOI is to make a detailed proposal for this.

7.5 Metadata records and DOIs

It was noted above that a metadata record might well be a tradeable item which is itself located by a DOI. There are reasons why this might be so for both level 1 and level 2 DOIs:

(1) Some early DOI users have experimented with assigning level 1 DOIs to *Intermediate objects* or *meta-objects* i.e. objects which are not themselves the end point of a specification by a unique identifier, but which may be a useful aggregation of information. Meta-objects are treated as if they were themselves a “first-class” object (i.e., assigned a separate identifier). The DOI is not be regarded as an identifier of a work, or performance or manifestation, but as an identifier of an

access route to information about or a service relating to a piece of intellectual property (just as my telephone number doesn't identify me, but it does identify a way of contacting me (David Martin)). The intermediate object may be a "shop front" or "parts list" for a source of the object in various formats from a particular vendor. Technically, intermediate objects are no more difficult to specify than other objects, but the business issues associated with them (such as who has rights to specify to or from what) may be complex. (Meta-objects may serve a useful purpose as a linking object specifying pointers to several related objects, and thus form an essential component of some identifier schemes).

(2) [DOI1] subsequently defined the scope of the DOI as follows: "1. Assign DOIs to Creations, but not to all Resources or Digital Objects, and not to Services; 2. Assign DOIs to Objects and to non-Digital Creations: Works and Packages..." [DOI paper 1, section 7]. We are now considering the allocation of DOIs to such Creations, and the holding of a metadata record in some way associated with the Creation. DOIs can be assigned to non-digital pieces of content. This raises questions about assigning metadata and determining the nature of the object identified.

An attractive way forward is to use the mechanisms of multiple resolution etc to enable "DOI Methods" or "DOI Options" such as metadata@DOI10.1000/2343 (a concept voiced at an early stage of DOI discussions). It seems that the URN framework would recognise DOI10.1000/2343 and metadata@DOI10.1000/2343 as two separate URNs (Leslie Daigle, pers.comm.). We need to examine closely how a record of DOI metadata is itself identified. This is a further item for investigation. The use of a syntax openly declaring Type (see section 5.2.6) in a level 2 DOI, as above, could also allow a separate metadata record to be identified e.g. as DOI10.1000/M/45678

8. Objections to the suggested approach

In working on the metadata implications of the DOI over the past few months, we have heard some objections on several occasions; these common "FAQs" are discussed briefly here.

8.1 "Isn't it better to start small?"

Can we not start with one application first and then build others without worrying about an overall data model?" The answer is that separate initiatives will never meet. The example of IT development is salient. Processes such as Rapid Application Development and Joint Application Development (RAD/JAD) with iterative prototyping are effective especially for small user-led developments. But if the applications so developed are to be interoperable they must conform to a corporate application architecture or data model. Otherwise they become silos, like digging a tunnel from two ends without a common plan. The data model does not have to be designed to perfection; it does not need to take months to complete; it just has to be logical and extensible; and it has to exist.

In a "trial and error" model of development from one application, without an underlying model, the DOI metadatabase must become a monolithic agglomeration of all likely metadata and is

likely to become a Frankenstein's monster; further there is no way to know for a given DOI object what sort of entity it is or what associated metadata it has (except to query the database and filter out the many gaps). The intention behind this suggested approach was well-meant: should the Foundation move quickly to develop something that works now in order to establish itself with a successful launch within one applications area, then move to other areas, which may have different requirements, adapting the system as it goes? Or should it move carefully forward to build, upfront, an infrastructure that ensures long term success for all applications, but by doing so, risk losing everything (because by the time it's built, everyone has moved on)? In fact the adaptations "as it goes" are likely in fact to be nothing more than a series of silo developments with little interoperability (and silos within a huge amorphous database). This is precisely the problem which has now been encountered in the Dublin Core approach, discussed extensively at the recent DC-6 meeting., and resulting in a significant and painful reappraisal of Dublin Core's future path.

The proposed approach, taking an existing developed, though still provisional, model and element set, does allow DOI to "start small" by concentrating on one or two particular DOI Genres and applications, but in the security that we have already staked out much of the wider ground within which the "small start" can be well-positioned and well-formed.

8.2 "Why discourage diversity?"

The suggested approach avoids diversity in the sense of preventing a Tower of Babel, with no interoperability. But it encourages diversity by being fully and logically extensible for specific local uses whilst retaining the guarantee of interoperability provided by a key set of core metadata in application-neutral form which can be used in any potential applications.

8.3 "It will be too much work for owners"

There is no escape from the fact that a move to electronic transactions requires information providers to become more rigorous in their approach to their internal data, and thus metadata (see e.g. [Neil]: "metadata; the matchmaker of the web"; [Dyck]: "Smooth data integration requires careful planning"; [Erickson]: metadata- the lifeblood of commerce systems"; [MDC]: the need for a defined "metadata interchange specification"). So some work will be necessary. There are some valid concerns on this score [Davidson] but deciding that it is going to involve effort and therefore cannot be done is not a solution, and will not make the problem go away; we can however minimise it. On the basis of "A stitch in time saves nine", by providing a careful and controlled approach we can minimise the amount of duplication work necessary. By insisting on controlled vocabularies we lay the basis for automation (for example, the proposed key elements *type*, *origination*, are automatable from short lists of options; the elements *title*, *creator* and other optional fields such as *relation* are clearly derivable from standard information typically held in workflow systems. Data mapping as shown in the attached Example 1 shows that it will be possible to generate most or all DOI metadata from other compatible schemas mapped through INDECS (this should include "legacy" standards such as MARC). Perhaps most important, when

a namespace registry system is in place, web tools such as RDF and XML themselves will provide the basis for highly automated metadata authoring solutions. The digital medium provides solutions as well as problems.

8.4 “Lets wait and see”.

It is our contention that we have enough experience to know what will happen. Without a disciplined metadata infrastructure DOI will not realise its full value.

9. Long term implications of DOI metadata: an infrastructure for content in a digital environment

This section expands on the need for structured metadata in the context of the overall aim of the DOI initiative. In particular, there are implications for the Registration Agencies for DOIs, and a change from the hitherto fairly relaxed rules about DOI application to a more enforced structured approach.

DOI is conceived as providing an identifier - and inescapably therefore a metadata framework - for multi-media transactions (E-commerce). Metadata has been described as “the lifeblood of e-commerce systems” [Erickson]. An object will contain different creations - audio, text, visual etc. The description of the object is a description of these component relationships. This is true for any creation (not just obvious multimedia); a book manifests one or more works, including illustrations and photographs, each of which have their own identities, rights owners and relevant agreements (licences). Much of this goes undocumented at present within publisher systems but will cause problems in the long-term.

In the digital environment all these component creations can be assembled, reassembled and modified to an unprecedented degree, and without a secure identification system for the creations, their creators and owners, the current problems faced by publishers in "physical" rights administration will be multiplied exponentially. We will not be able to licence, sell or protect our property effectively without participating more broadly in a standardised "network" of multi-media identifiers. This realisation was behind the conception of the DOI.

We are seeing the beginnings of the development of such a distributed "identifier network" of creation identifier strings, linking all kinds of owners and users. That is the importance of interoperability of identifiers. The CISAC CIS plan is working towards securing one particular part of this network. Other industries (banking, credit cards, airlines) have not dissimilar global "identifier networks", but for creations the elements are more diverse and more complex, and the chains longer. DOI promises to provide a vital piece of the identifier network: identifying digital objects, usable alongside ISBNs, ISRCs, ISWCs, ISANs and the rest for a whole new and massive range of creations; and even more promising, DOI can act as a "substitute" identifier for non-digital Creations (packages, performance and works). So rights management data might be made up of a string of DOIs, linking at the end to parties and agreements.

If the systems used to process these (growing) millions of data-strings have no way of knowing which part of the string (if any) a DOI is supposed to represent; of what it actually refers to; and of whether the person who linked it to the other identifiers had any idea what he was doing; then we will encounter significant problems in making the DOI system work. At present rights administration is hampered by the inconsistent and sometimes incorrect use of established identifiers like ISBN, ISRC and UPC/EAN. Sometimes this causes serious problems. A dumb, *unregulated* DOI, however, promises an undreamt of level of unreliability and confusion, and so serious rights administrators (which include publishers, libraries and corporate users as well as copyright societies and clearing houses) will eventually just ignore it and find other solutions. Note that every transaction, including accessing a cited document, involves rights administration, even in those cases where all rights are waived or material is in the public domain. Worthwhile systems will have to positively recognise the deliberately unprotected as well as the protected (e.g. the CISAC community has experienced this recently in recognising that public domain works such as Beethoven symphonies must be given ISWCs so that everyone can positively identify that no clearance is needed).

If the DOI identifier is dumb, systems have to refer to metadata to determine what it is dealing with; that metadata has to be highly structured, as the creations which are transacted and managed may have been manifest in the widest variety of other creations. The metadata used for (say) academic citations is going to be accessed by systems dealing with (say) computer games, so it must be easily intelligible, which means it must be highly standardised and genre- and sector-neutral. Rights administration is used as an example, but this applies to any automated process which is crunching through stacks of identifiers for whatever purpose (data-mining or policing, for example). The main point of DOI in the long-run is to support predominantly automated e-commerce and librarianship (these terms used in their widest sense). If systems are going to draw blanks on five out of every ten DOIs processed because either (a) software hasn't been told of new metadata varieties, or (b) the allocator of the DOI has not registered data, or done so incorrectly or (c) the use of the DOI is inappropriate or ambiguous, then DOI is useless.

This means that the metadata structure must be stable and highly standardised, and at the same time (because DOI applies to any creation type) consistent with all the existing metadata schemas attached to major identifiers. Such a thing is quite possible. XML and RDF provide the basis for such a framework. The INDECS project in which DOI is participating is designed to develop precisely such a composite standard. The top-down, abstract-to-specific approach is the only one that can work as a logic-driven digital tool for the huge range of potential coverage of DOI.

If we agree some basic standard metadata elements and structure, it must be controlled (validated) at the point of registration. That will carry a cost; the cost of adding value to the system, which if well done will be recouped by an effective DOI system which does far more than any loose "resource discovery" tool possibly can. We expect a DOI in the long run it to be able to do many different things for many different people, far more diverse and complex than credit cards. We will not achieve this without very tight up front control and structure. No-one will trust it, so they won't use it.

10. Implications

The DOI has been conceived on a very large scale (even though its originators did not have that intention): its scope is any piece of intellectual content (in any format), providing digital services relating identified entities to E-commerce (including in “commerce” free transactions). In the digital environment, these services need to be interoperable. Therefore we are creating an infrastructure which potentially is applicable to almost anything in which rights pertain; we may start with journal articles but by definition we have conceptually included e.g. a motor car (although pragmatic considerations may prevent that assignation, other limits are more easily tested). Further it links to any conceivable transaction service. It is an identifier which has broad scope (the nearest similar being UPC/EAN product codes and credit card numbers): the consequence is it becomes “all things to all people”.

Therefore the potential scale of the DOI development is large: conceptually bigger than existing systems such as EAN/UPC and CIS, both of which have been multi-year multi-organization efforts; bigger than conceptual schema such as URNs which have been under development for several years. The development will be time-consuming and complex, and will need good project management. There is one way to focus the effort and that is to limit the IDF’s development to building an infrastructure, not all the applications.

The member organizations of the Foundation have been instrumental in working on some of the issues to date, but to take the DOI forward it is now appropriate to move to a more structured process. We have begun this with the proposal for task groups (each member organization represented on at least one); a further suggestion is that a technical steering board might consist of the chairs of these task groups, to coordinate developments and timetables. This would be separate from the elected Board and might necessitate much more active involvement in detailed issues. The model of CIS might be appropriate.

11. Summary: proposed DOI metadata development strategy

The DOI Foundation proposes that its metadata be “well structured” as defined here. It proposes to use the INDECS (Interoperability of Data in E Commerce Systems) initiative, part of the European Commission’s Info 2000 activity, as a forum for using the “well-formed” metadata approach, and arriving at a consensus on this issue (IDF is a participating member organization of INDECS). Among many advantages of this is the fact that INDECS is being run by individuals with experience of constructing commercial intellectual property metadata systems which are working. The IDF will budget for appropriate expenditure in conjunction with INDECS, rather than a separate metadata activity. IDF proposes that DOI metadata should be expressible in RDF and will prepare an RDF schema for the metadata.

It must be noted that several of the concepts relevant to this discussion are themselves only now being finalised or are entering the standards processes (e.g. DC, RDF, XML).

There are then several clear challenges to be solved as our next steps:

- Agree a metadata framework, proposed here in draft form, which has at the top level a schema such as INDECS (hopefully, INDECS/Dublin Core consensus) for generic data elements, and recognises a wide range of compatible systems for lower level qualification.
- To define how this (mandatory) kernel metadata is to be injected into the DOI process. RDF is fine as far as it goes, as a means of expressing metadata, but it is just a language (or a dialect of a language anyway). Where is the metadata to be stored, found, and referenced?
- To define metadata appropriately and accessibly so that a means of finding out the nature of both the entity identified and its related creations can be provided.

As specific tasks arising from this paper:

- Encourage testing and review of the proposed DOI element set, the kernel set and the basis for extensions
- Define an initial set of DOI Genres
- Work with one or more organisations in a genre to define extensions to the kernel and one or more demonstration applications using that metadata in that Genre (e.g: STM journals, reference linking)
- Propose an appropriate syntax for level 2 DOIs
- Consider appropriate mechanisms for declaration of metadata (RDF etc) and initiate work on corresponding standards
- Consider necessary standards for an "authority" record for each metadata declaration.

References

- Baker Baker, Thomas
 Languages for Dublin Core
 D-Lib magazine, December 1998
 <http://www.dlib.org/dlib/december98/12baker.html>
- Bearman Reconciling Metadata Requirements form the Dublin Core and INDECS/DOI
 Communities
 Bearman, D., Miller, E., Rust, G., Trant, J. and Weibel, S.
 D-Lib magazine, January 1999?, in preparation; reference to be added
- Bryan Bryan, Martin
 An Introduction to the Extensible Markup Language (XML)
 ASIS Bulletin, Oct/Nov 1998
 <http://www.asis.org/Bulletin/Oct-98/bryanart.html>
- CIMI Consortium for the Computer Interchange of Museum Information
 http://www.cimi.org/downloads/CIMI_profile/profile.htm
- CIS Hill, Keith (1997).
 CIS - A Collective Solution for Copyright Management in the Digital Age
 Copyright World, issue 76, 1997, pp18-25
- Cronin Cronin, B et al
 Invoked on the Web
 Journal of the American Society of Information Science 49(14): 1319-1328
 (1998)
- Davidson Digital Object Identifiers: Promise and Problems for Scholarly Publishing
 Davidson, L.A. and Douglas, K.
 Journal of Electronic Publishing Volume 4, Issue 2; December, 1998
 <http://www.press.umich.edu/jep/04-02/davidson.html>
- DOI1 The Digital Object Identifier initiative: current position and view forward; DOI
 discussion paper number 1
 DOI 10.1000/92 (<http://dx.doi.org/10.1000/92>)
- DOI3 The Digital Object Identifier: creating services using DOIs.
 DOI discussion paper number 3 (in preparation; further information available
 from the Foundation)
- DOI4 The Digital Object Identifier: using DOIs in a local environment.
 DOI discussion paper number 4 (in preparation; further information available

from the Foundation)

- DOIDiscuss DOI discussion list mail archive
<http://www.doi.org/mail-archive/discuss-doi/maillist.html>
- DOISyntax DOI syntax discussion list mail archive
<http://www.doi.org/mail-archive/syntax/maillist.html>
- DublinCore Dublin Core Metadata
http://purl.oclc.org/metadata/dublin_core/
- Dyck Smooth data integration requires careful planning
Timothy Dyck
PC Week 02.02.98
<http://www.zdnet.com/pcweek/sr/0202/02data.html>
- Erickson Erickson, John (1998)
Metadata Initiatives and the DOI: Implications for Electronic Commerce and
Copyright Management Services
Dialogue, Summer 1998 Number 8
<http://www.ybp.com/yrm/trialogue/898/898metad.htm>
- FGDC Federal Geographic Data Committee
<http://www.census.gov/geo/www/standards/scdd/CDsupplement.html>
- Ford/Harter Ford, Charlotte E., & Harter, Stephen P. (1998)
The Downside of Scholarly Electronic Publishing: Problems in Accessing
Electronic Journals through Online Directories and Catalogs
College and Research Libraries, 59, 335-346
- Hitchcock Hitchcock, S et al (1998)
Linking electronic journals: Lessons from the Open Journal project
D-Lib magazine December 1998
<http://www.dlib.org/dlib/december98/12hitchcock.html>
- Hunter Hunter, K. (1998)
Adding Value by Adding Links.
Journal of Electronic Publishing, Vol. 3, Issue 3, March
<http://www.press.umich.edu/jep/03-03/hunter.html>
- IDF Dirrep 4 Directors Report to members of the International DOI Foundation, Number 4,
1 October 1998
Available to IDF members only at <http://www.doi.org/idf-members/index.html>

- IFLA IFLA Study Group on the Functional Requirements for Bibliographic Records (1998).
Functional Requirements for Bibliographic Records.
Final report, K.G. Saur Munchen (1998)
Also available at <http://www.nlc-bnc.ca/ifla/VII/s13/projects.htm>
- INDECS Interoperability of Data in E Commerce Systems
European Commission, DG XIII, Info 2000 programme
<http://www.indecs.org>
- ISO 11179 International Standard ISO 11179, Information Technology-
Specification and Standardization of Data Elements
<http://www.sdct.itl.nist.gov/~ftp/18/other/Standards/iso11179/>
- Jul Jul, Erik
InterCat and NetFirst provide access to the World Wide Web
OCLC Newsletter 225
<http://www.oclc.org/oclc/new/n225/mn.htm#inc>
- Lynch Lynch, Clifford
Identifiers and Their Role In Networked Information Applications
ASIS Bulletin, December 1997
<http://www.asis.org/Bulletin/Dec-97/lynch.htm>
- Martin Martin, James (1982)
Strategic Data Planning. Prentice-Hall
- MDC Metadata Coalition: The Meta Data Interchange Specification Initiative
<http://www.mdcinfo.com/papers/intro97.html>
- Metadata Digital Libraries: Metadata Resources
<http://www.ifla.org/II/metadata.htm>
- Miller Miller, Eric (1998)
An Introduction to the Resource Description Framework
D Lib Magazine, May 1998:
<http://www.dlib.org/dlib/may98/miller/05miller.html>
also re-printed in ASIS Bulletin, Oct/Nov 1998:
<http://www.asis.org/Bulletin/Oct-98/ericmill.html>
- Neil Metadata: The matchmaker of the Web
Stephanie Neil
PC Week
<http://www.zdnet.com/pcweek/news/0504/04meta.html>

| | |
|------------|---|
| Newton | Using levels of abstraction to name data elements Judith J. Newton http://www.tdan.com/i007ht01.htm |
| NISO-DOI | (Draft) American National Standard Digital Object Identifier ANSI/NISO Z39xx-199x http://www.niso.org (In preparation) |
| NIST | www.nist.gov/L8 |
| Paskin | Paskin, Norman (1999) Towards Unique Identifiers submitted for Proceedings of the IEEE, Special Issue on "Identification and Protection of Multimedia Information" (Copies available from author on request) |
| RDF | W3C: Resource Description Framework http://www.w3.org/RDF/ |
| RDFmod | Resource Description Framework (RDF) Model and Syntax Specification http://w3.org/TR/1999/PR-rdf-syntax-19990105 |
| Roszkowski | Michael Roszkowski and Christopher Lukas A Distributed Architecture for Resource Discovery Using Metadata D Lib Magazine, June 1998 http://www.dlib.org/dlib/june98/scout/06roszkowski.html |
| Rust | Rust, Godfrey (1998) Metadata: The Right Approach. An Integrated Model for Descriptive and Rights Metadata in E-commerce D Lib magazine, July 1998 http://www.dlib.org/dlib/july98/rust/07rust.html |
| StPierre | St. Pierre, Margaret & La Plant, William (1998) Issues in Crosswalking Content Metadata Standards http://www.niso.org/crosswalk.html |
| USMARC | MARC Standards (1998) Library of Congress Network Development and MARC Standards Office http://www.loc.gov/marc |

APPENDIX: Examples of DOI metadata following the basis set out in the Discussion paper

(author: Godfrey Rust, INDECS/Data Definitions)

Please note: suggested possible extensions are highly speculative and used for illustration rather than definition. For DOI and INDECS purposes many more examples will be tested and requests for illustrations of specific items (including particularly difficult examples!) are welcomed.

Example 1: Book

- DOI “key” or kernel metadata
- Possible extended kernel for this DOIGenre
- Fullrecord

Example 2: Music CD

- Kernel
- Possible extended kernel for this DOIGenre

Example 3: Poem

- Kernel
- Possible extended kernel for this DOIGenre

Example 4: Journal issue

- Kernel
- Possible extended kernel for this DOIGenre

Example 5: Journal article

- Kernel
- Possible extended kernel for this DOIGenre

DOI Kernel Metadata Example 1

THE SUBTLE KNIFE Novel

(including mapping from using Title Information Project elements)

Qualifiers in Roman type are from DOI namespace. Qualifiers in italics are from other namespace. Shaded elements are those added from previous

| | <i>DOI</i> kernel Elements.qualifier | Value | Namespace |
|----|---|-----------------------------------|--------------------|
| 1 | DOI | 10.0000/0001 | TIP 0102-[DOI] |
| 2 | DOIGenre | Book | TIP 1401 |
| 3 | Identifier.ISBN | 0590112899 | TIP 0102-[IB] |
| 4 | Title | The Subtle Knife | TIP 0301 |
| 6 | Primary Type | Physical Manifestation | Der. TIP 1401 |
| 7 | Origination | Edition | Der. TIP 0803 |
| 13 | Event.Creation | Agent=Philip Pullman, Role=Author | TIP 0504, TIP 0401 |

Possible kernel metadata extension for Book/Physical Manifestation

Form/Physical Format

Form/Textual Genre

Form/Language

Event/Type=First Publication

Possible Extended Kernel Metadata for this Book/Physical Manifestation

(NB: der = derived systematically from)

| | <i>DOI</i> kernel Elements.qualifier | Value | Namespace |
|----|---|---|--|
| 1 | DOI | 10.0000/0001 | TIP 0102-[DOI] |
| 2 | DOIGenre | Book | TIP 1401 |
| 3 | Identifier. <i>ISBN</i> | 0590112899 | TIP 0102-[IB] |
| 4 | Title | The Subtle Knife | TIP 0301 |
| 6 | Primary Type | Physical Manifestation | Der TIP 1401 |
| 7 | Origination | Edition | Der TIP 0803 |
| 9 | Form.Physical | Paperback | TIP 1402 |
| 9 | Form.Textual Genre | Novel | TIP 1301 |
| 9 | Form.Language | English | TIP 0902 |
| 13 | Event.Creation | Agent=Philip Pullman, Role=Author | TIP 0504, TIP 0401 |
| 13 | Event.Dissemination. <i>First Published</i> | Agent=Scholastic Childrens Books, Role=Publisher; Agent=Point, Role=Imprint Name; Time=1998 | TIP 0604, TIP 1704; TIP 1703; TIP 1801 |

DOI Metadata Example 1 continued
THE SUBTLE KNIFE Novel

Full DOI Metadata record for this Book/Physical Manifestation including non-kernel elements

| | DOI Elements.qualifier | Value | Namespace |
|----|--|---|----------------------|
| 1 | DOI | 10.0000/0001 | TIP 0102.[DOI] |
| 2 | DOIGenre | Book | TIP 1401 |
| 3 | Identifier. <i>ISBN</i> | 0590112899 | TIP 0102.[IB] |
| 3 | Identifier. <i>EAN13</i> | 97800590112895 | TIP 0102.[EN] |
| 4 | Title | The Subtle Knife | TIP 0301 |
| 4 | Title. <i>Series</i> | His Dark Materials | TIP 0301 |
| 5 | Description. <i>Short Review Quote</i> | “The most ambitious work since Lord of the Rings...as intellectually thrilling as it is magnificently written”, Amanda Craig, New Statesman | TIP 1001.[03] |
| 6 | Primary Type | Physical Manifestation | Der TIP 1401 |
| 7 | Origination | Edition | Der TIP 0803 |
| 8 | Mode | Visual | Der TIP 1401 |
| 9 | Form.Textual Genre | Novel | TIP 1301 |
| 9 | Form.Physical | Paperback | TIP 1402 |
| 9 | Form.Physical. <i>Presentation</i> | No dust jacket | TIP 1403-0 |
| 9 | Form.Physical. <i>Binding</i> | Massmarket Paperback | TIP 1403-1 |
| 9 | Form.Language | English | TIP 0902 |
| 9 | Form.Completion. <i>Publication Status</i> | New Edition | TIP 1701 |
| 10 | Context.Audience | General | TIP 1101 |
| 11 | Subject.Topic | Science Fiction | [Dewey/LC] |
| 11 | Subject.Thing | Aurora Borealis | [Dewey/LC] |
| 12 | Extent.Dimension. <i>Height</i> | Number=170, Measure=MM | TIP 1502, 1503, 1504 |
| 12 | Extent.Dimension. <i>Width</i> | Number=88, Measure=MM | TIP 1502, 1503, 1504 |
| 12 | Extent.Dimension. <i>Number of Items</i> | Number=1, Measure=Unit | TIP 1505, 1503, 1507 |
| 12 | Extent.Dimension. <i>Number of Pages</i> | Number=342, Measure=Unit | TIP 1501 |
| 13 | Event.Creation | Agent=Philip Pullman, Role=Author | TIP 0504, TIP 0401 |

| | | | |
|----|--------------------------------------|---|--|
| 13 | Event.Creation | Agent=TW Typesetting, Role=Typesetter | TIP 0604 |
| 13 | Event.Creation | Agent=Cox & Wyman, Role=Printer | TIP 0604 |
| 13 | Event.Creation | Agent=Black Sheep, Role=Designer, Component=Cover | TIP 0604 |
| 13 | Event.Dissemination | Agent=Scholastic Childrens Books, Role=Publisher; Agent=Point, Role=Imprint Name; Time=1998 | TIP 0604, TIP 1704; TIP 1703; TIP 1801 |
| 14 | CreationLink.Version. <i>Edition</i> | Link Creation Identifier.ISBN=0590112222, Sequence=2 | TIP 0803 |
| 14 | Creation Link.Component.Series | Link Creation Identifier.DOI=10.0000/0007, Sequence=2 | |

DOI Kernel Metadata Example 2

Paul Simon/GRACELAND CD

| | DOI Elements.qualified | kernel Value |
|----|-----------------------------------|--|
| 1 | DOI | 10.0000/0002 |
| 2 | DOIGenre | Audio |
| 3 | Identifier. <i>UPC</i> | 7599254472 |
| 4 | Title | Graceland |
| 6 | Primary Type | Physical Manifestation |
| 7 | Origination | Original |
| 13 | Event.Creation | Agent=Paul Simon, Role=Featured Artist |

Possible kernel metadata extension for Audio/Physical Manifestation:

Form/Physical Format

Form/Digital Format

Form/Audio Genre

Extent/Number/Tracks

Event/Type=First Release

Possible Extended Kernel Metadata for this Audio/Physical Manifestation

| | DOI Elements.qualified | kernel Value |
|----|---|---|
| 1 | DOI | 10.0000/0002 |
| 2 | DOIGenre | Audio |
| 3 | Identifier. <i>UPC</i> | 7599254472 |
| 4 | Title | Graceland |
| 6 | Primary Type | Physical Manifestation |
| 7 | Origination | Original |
| 9 | Physical Format | CD |
| 9 | Digital Format | ADD |
| 9 | Audio Genre | Album |
| 12 | Extent.No of Tracks | Number=12, Measure=Units |
| 13 | Event.Creation | Creation, Agent=Paul Simon, Role=Featured Artist |
| 13 | Event.Dissemination. First Published | Agent=Warner Bros, Role=Record Company, Time=1986 |

DOI Kernel Metadata Example 3

Wendy Cope/MEN AND THEIR BORING ARGUMENTS Poem

| | <i>DOI</i> kernel Elements | Value |
|----|-----------------------------------|---|
| 1 | DOI | 10.0000/0003 |
| 2 | DOIGenre | Textual Work |
| 3 | Identifier. <i>ISWC-L</i> | L123456789 |
| 4 | Title | Men And Their Boring Arguments |
| 6 | Primary Type | Work |
| 7 | Origination | Original |
| 13 | Event | Creation: Agent=Wendy Cope, Role=Author |

Possible kernel metadata extension for Textual Work

Origination/if Version=Creation Link to source

Form/Textual Genre

Possible Extended Kernel Metadata for this Textual Work

| | <i>DOI</i> kernel Elements | <u>Value</u> |
|----|-----------------------------------|---|
| 1 | DOI | 10.0000/0003 |
| 2 | DOIGenre | Textual Work |
| 3 | Identifier. <i>ISWC-L</i> | L123456789 |
| 4 | Title | Men And Their Boring Arguments |
| 6 | Primary Type | Work |
| 7 | Origination | Original |
| 9 | Textual Genre | Poem |
| 13 | Event | Creation: Agent=Wendy Cope, Role=Author |

DOI Kernel Metadata Example 4

D-Lib, January 1999

| | DOI kernel Elements | Value |
|----|----------------------------|----------------------------------|
| 1 | DOI | 10.0000/dlib/january99 |
| 2 | DOIGenre | Journal |
| 3 | Identifier.ISSN | 1082-9873 |
| 4 | Title | D-Lib, January 1999 |
| 6 | Primary Type | Digital Manifestation |
| 7 | Origination | Original, Compilation |
| 13 | Event.Creation | Agent=Williams Arms, Role=Editor |

Possible kernel metadata extension for Journal+Digital Manifestation

Title/Issue Title=

Identifier/Volume number=

Identifier/Issue number=

Form/Journal Frequency= Form/Digital Format=html

Extent/Number=component articles

Event/Type=Dissemination/Role=Publisher/Time=Date of Publication

Possible Extended Kernel Metadata for this Journal

| | DOI kernel Elements | Value |
|----|----------------------------|---|
| 1 | DOI | 10.0000/dlib/january99 |
| 2 | DOIGenre | Journal |
| 3 | Identifier.ISSN | 1082-9873 |
| 3 | Identifier.Original url | http://www.dlib.org/dlib/january99 |
| 3 | Identifier.Volume | 5 |
| 3 | Identifier.Issue | 1 |
| 4 | Title | D-Lib |
| 4 | Title.Issue | January 1999 |
| 6 | Primary Type | Digital Manifestation |
| 7 | Origination | Original, Compilation |
| 9 | Frequency | Monthly |
| 12 | Extent.Components | Number=6, Measure=Units |
| 13 | Event.Creation | Agent=Williams Arms, Role=Editor |
| 13 | Event.Dissemination | Publisher=CNRI, Time=19990115 |

DOI Kernel Metadata Example 5
Journal Article, D-Lib, January 1999

| | DOI kernel Elements | Value |
|----|----------------------------|---|
| 1 | DOI | 10.0000/0005 |
| 2 | DOIGenre | Textual Work |
| 3 | Identifier. <i>PII</i> | 99999999 |
| 4 | Title | Geographic Names: The Implementation of a Gazetteer in a Georeferenced Digital Library |
| 6 | Primary Type | Work |
| 7 | Origination | Original |
| 13 | Event.Creation | Agent=Linda L Hill, Agent=James Frew, Agent=Qi Zheng; Role=Author; Affiliated Agent=Alexandria Digital Library Project, University of California, Place=Santa Barbara |

Possible kernel metadata extension for Textual Work

Origination/if Version=Creation Link to source

Form/Textual Genre

Possible kernel metadata extension for Textual Work = Journal article

Creation Link=First publication

| | DOI kernel Elements | Value |
|----|--|---|
| 1 | DOI | 10.0000/0005 |
| 2 | DOIGenre | Textual Work |
| 3 | Identifier. <i>PII</i> | 99999999 |
| 4 | Title | Geographic Names: The Implementation of a Gazetteer in a Georeferenced Digital Library |
| 6 | Primary Type | Work |
| 7 | Origination | Original |
| 9 | Form.Textual Genre | Journal article |
| 13 | Event.Creation | Agent=Linda L Hill, Agent=James Frew, Agent=Qi Zheng; Role=Author; Agent Affiliate*=Alexandria Digital Library Project, University of California; Place=Santa Barbara, CA |
| 14 | Creation Link. Component.First Publication | Link Creation Identifier.DOI=10.0000/dlib/january99, Sequence=2 |

*This example introduces a possible technique (not described in the paper) of showing the institutional affiliation of authors (in Event.Creation), using an “Agent Affiliate” element. The INDECS schema shows that this information properly belongs to Person metadata – in this case, the Alexandra Digital Library Project and the University Of California are both legal Persons with

whom the authors have a relationship at this time. However, this affiliation can be included within the nesting of elements in the Event to directly associate the institutions with the creation. Such “intentional de-normalisation” of the metadata model can be well handled in an XML DTD, allowing the data to be reconstructed into its correct place in the model if required.